INTEGRATION OF DISTRIBUTED HEALTHCARE INFORMATION SYSTEMS: APPLICATION OF CEN/TC251 ENV13606

J.A. Maldonado[*], M. Robles, C. Cano

E.U.I., Technical University of Valencia, Valencia 46022, Spain.
*Email: jamaldo@upvnet.upv.es

Abstract-**Everyday, more and more, the healthcare sector requires the exchange of healthcare information between professionals from different disciplines and institutions. To support co-operative work among health professionals and institutions it is necessary to share healthcare information about patients in a meaningful way. But, nowadays, in most hospitals the heath data is distributed across several heterogeneous and autonomous information systems whose interconnection is difficult to achieve. Integration of such system may bring about many advantages such as consistent patient health records or interdepartmental workflows. In this paper, an overview of a computing system prototype, currently under development using CORBA and Java, is presented. It allows healthcare professionals to access patient information stored in heterogeneous autonomous information systems through a set of formal aggregates of health data based on the healthcare record architecture ENV13606 from CEN/TC251.**
Keywords- **systems integration, medical records, electronic healthcare record architecture, CORBA.**

I. INTRODUCTION

A common scene within most hospitals nowadays is the distribution of health data along departamental information systems. This leads to fragmented and heterogeneous data resources and services, all of them containing health data about patients, and contributes to the emergence of the so-called *islands of information*. The main reasons for the existence of information system heterogeneity in hospitals are:

- The complexity and great variety of healthcare actions and protocols, the diversity of organisations not only regarding structure or size but also political, economical or cultural aspects and the preferences of health professionals' groups makes it very difficult to develop a single computer system that could effectively serve the information needs of an entire hospital. As a result, most hospitals have developed their information systems on a departament-oriented basis.
- The fragmentation of the health IT market, where a great variety of specialised products exists, whose interconnection is usually difficult to achieve. On the other hand the products which try to cover the full functionality of the entire Hospital Information System (HIS) often lack specialised functionality.

- Legacy information systems, which often are very old and significantly resist modification and evolution.

As a result the health data are organised and managed by several heterogeneous and autonomous information systems. Heterogeneity in distributed information systems can be present at different levels [1]: hardware that supports the database management system, operating systems, communication protocols and in the database systems that hold the data which can be further divided into:

- those due to differences in the database management system which can be based on different data models (relational, semantic, functional, hierarchical, network, object-oriented, etc.) and/or based on the same model but from different vendors (Oracle, Informix, Sybase, etc, which are relational DBMS developers).
- those due to differences in the semantics of data.

Nowadays, the healthcare sector is undergoing a change. The traditional single doctor-patient relationship is being replaced by one in which a team of healthcare professionals from different disciplines and institutions is responsible for patient health. This new context requires a high level of interoperability and data sharing among professionals and institutions involved in the healthcare of a patient. This crucially depends on an ability to exchange information about patients while preserving its original meaning. In the absence of this information, tests may be repeated or prior findings ignored, and in emergency care lifesaving information may be unavailable. Briefly, what is required is that everyone involved in the delivery of healthcare to a patient should be able to access all the relevant patient's healthcare information. Unfortunately, data sharing is often hindered by the fact that the data are distributed among several autonomous and heterogeneous (database) systems.

II. INTEGRATION OF DISTRIBUTED HEALTHCARE INFORMATION

The two main approaches that enable organisations to integrate their heterogeneous and autonomous information systems are the revolutionary and evolutionary approaches [2]. The former consists of acquiring a new system, that has

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle	Contract Number
Integration of Distributed Healthcare Information Systems: Application of CEN/TC251 ENV13606	**Grant Number**
	Program Element Number

Author(s)	Project Number
	Task Number
	Work Unit Number

Performing Organization Name(s) and Address(es)	Performing Organization Report Number
E.U.I, Technical University of Valencia Valencia 46022, Spain	

Sponsoring/Monitoring Agency Name(s) and Address(es)	Sponsor/Monitor's Acronym(s)
US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Report Number(s)**

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom. , The original document contains color images.

Abstract

Subject Terms

Report Classification	Classification of this page
unclassified	unclassified

Classification of Abstract	Limitation of Abstract
unclassified	UU

Number of Pages
4

the potential to embrace all the functions required by the organisation, and discard the existing ones. This approach has some drawbacks: no single system can optimally support all user roles and no single system can be populated by all the necessary data. The domain specific systems generally completely fulfil the information requirements of a particular department or unit and the need to protect the investments in existing system. The latter consists of achieving some kind of global and uniform treatment of the data maintaining at the same time the autonomy of the underlying databases. The main disadvantage of this approach is its complexity since it requires a deep knowledge of the semantics of the data [3].

There are two different ways of system integration according to its scope: a) integration of systems within a company or organisation to allow the direct and uniform access to all relevant data and b) global connection of databases and information systems of different companies or organisations (global information-sharing systems). It is highly desirable that the technological solution used for intra-organisation communication or integration should be compatible (ideally the same) to the one used for inter-organisation communication.

On the other hand, the Electronic Healthcare Record (EHCR) is on the way (although the more we get involved in it the further away it seems to be) and anyone starting a project to integrate Health Information Systems at this time, should consider the implications of the EHCR and whether their integration strategy will be compatible with and support the emerging EHCR.

Any integration solution has to define a way to represent the patients specific health data in such a way that their original meaning is preserved through faithful preservation of content and context. Standardisation of EHCR architecture is vital if the clinical information is to be transferred outside the organisation/department where it was created. Much work has been done in the field of EHCRA standardisation, in Europe several groups have developed their own architecture. In particular, Work group I of CEN/TC251 (European Committee of Normalisation, Technical Committee 251) has developed a pre-standard known as ENV 13606 [4]. This pre-standard defines a conceptual data model which is capable of structuring any medical data in a uniform way, presenting the multitude of different facts while the preserving meaning and context of the data.

III. DATA ENGINEERING

Our solution is based on specifying a generic computer system that lets health professionals retrieve, from the underlying health data repositories, the patient data that they need on-line and present the data in an integrated common way. The system manages a "virtual" health record which is assembled "on the fly" from data held in multiple (database) systems. This requires the ability to map the diverse data structures into a common one. In other words, the presentation of the different parts of the patient EHCR on the client in a integrated way implies the definition of an underlying object model in which the structure of a patient EHCR is defined, i.e. an EHCR architecture.

The EHCR architecture used in our project is ENV13606 from CEN/TC251. This architecture is used to provide the clients with a way to build electronic health records as well as a unified view of a patient's medical record. The architecture provides the structures to build "on the fly" a part of or the entire patient's healthcare record drawn from any number of heterogeneous databases systems, i.e. ENV13606 is used to define the retrievable objects.

The basic elements of a medical record, as defined by ENV 13606, are:

- The EHCR Root Architectural Component: represents the component which has as its content the subject of care's EHCR.
- Link items allow the definition of relationships between two Components in an EHCR.
- Selected Component Complexes (SCC) permit Record Components to be re-used by providing additional views of the original data.
- Data items represent an observation by an agent at a particular time and place. They constitute an aggregate of information that cannot safely be disaggregated. Several types are defined in the pre-standard.
- Original Component Complexes (OCC) serves the purpose of grouping Record Components. There are four types of such groupings:

  - Folder OCCs which represent whole sections of a subject's life-long health record.
  - Composition OCCs which hold a set of record components relating to one time and place of care delivery, a single session of recording or a single document included in the EHCR.
  - Headed Section OCCs represent a sub-division of data within a Composition OCC, whose contents have a common theme or are derived from a common healthcare process.
  - Cluster OCCs which group together a closely related set of Data Items.

Figure 1 describes the three levels of our system model: conceptual, semantic and data which will be described next.

As it was stated before, ENV13606 is used as a canonical electronic healthcare record architecture which defines how electronic records must be built. ENV13606 defines the components which are necessary to allow the content of a healthcare record to be constructed, used, shared and maintained. It represents the conceptual level of our system.

The components of ENV13606 have been defined at a high level of abstraction to provide a flexible model capable of representing any entry in a healthcare record, independently of the healthcare institution, speciality or professional. Thus, the ENV13606 classes can be easily extended to represent terms or concepts from the medical domain (e.g. GP Record, inpatient stay, discharge report, transfer, demographics, blood

pressure, protein S level, etc.). This representation leads to a set of new classes which extend (specialise) the ones defined in ENV 13606, we call them archetypes. For instance, a discharge report can be defined by using a Composition OCC which contains other terms (which are themselves represented by a ENV 13606 construct) such as patient details, diagnosis, comments, medication, etc. These sub-terms may be themselves based on others and so on. The data items are at the lowest level of the hierarchy and therefore are the basic blocks to construct other components. Examples of data items are: patient's name, main diagnosis code (drawn from an international classification of diseases), discharge date, discharge reason, etc. This allows a high level of reutilization, already defined archetypes may be used to build new ones. The archetypes are the core of our integration solution, their purpose is to make public the information content in the underlying databases and, at the same time, to hide technical details (heterogeneity) of the data repositories. They constitute a semantic layer over the underlying databases associating them with domain specific semantics, thus true integration is performed at a metal-level instead of at data level. One of the main advantages of this approach is that it allows each hospital unit or service to define its own view of the medical record.

Since the health data resides on the underlying databases, there should be defined some kind of mapping information relating the archetypes to the databases schemas. Currently, in our system the mapping information is expressed using SQL clauses. It is important to remark that it is not only necessary to map the information content but also any available context information in order to safeguard the original meaning of data. ENV 13606 describes which context information must accompany any piece of health information.
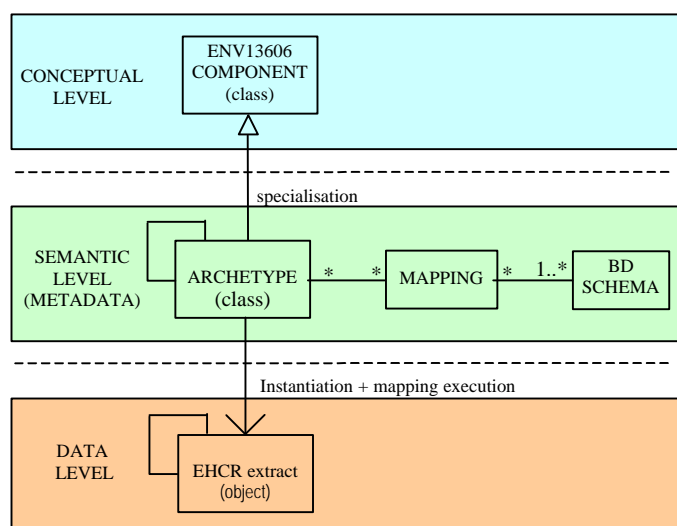

Fig. 1. System model. It consists of three levels: conceptual, semantic, which represent the system metadata, and data level which contains the actual health data for a patient.

The archetypes are also associated with presentation information to allow the definition of the appearance of then information when presented to the user. The archetypes and

their relationships, the archetypes presentation information, the schemata of the integrated databases and the mappings constitute the semantic level of our system.

The lowest level of our system is the data level which represents the actual health data about patients as one or more instances of any archetype. As stated before, each archetype contains the necessary information to create and populate (through mapping execution) all the objects that are necessary to construct the extract of the health record that it defines.

IV SYSTEM ARCHITECTURE AND IMPLEMENTATION

The starting points in the design of our prototype were to implement a system that allows the underlying databases to retain their autonomy and capable of adapting to a such a dynamic environment characterised by heterogeneous and autonomous information systems. The basic architecture is illustrated in Figure 2. It consists of a set of distributed CORBA objects (CORBA object are server by nature) implemented in Java. Several reason can be put up for the use of Java, it is a real portable and all-purpose object-oriented programming language which allows hardware and operating system independence, it comes with a complete set of very useful functionalities suitable for our project, such as access to databases through JDBC, generation of GUI''s, CORBA and internet technology. Concretely, the "marriage" Java-CORBA is very successful as both object-oriented technologies fit perfectly and complement each other.

The servers that make up the system are:

- The metadata server is the module which is in charge of managing the system metadata. It manages a database (implemented by using an object oriented database, ObjectStore for Java) which contains the archetypes definition, the underlying databases schemas, the archetypes-schema mappings, information about the location of patients' social-demographic data, general information about the underlying databases and network addresses. Two visual tools have been developed to assist in management of the metadata: the archetype editor and the schema manager. The former facilitates the edition of archetypes (it allows the creation of new ones from scratch or the reuse of existing ones), validates the correctness, allows the classification them into groups in order to ease the search, define the mappings with the component databases schemas and finally it controls the versioning (all prior versions are kept for legal reasons). The latter allows the automatic retrieval of schemas from the underlying databases (at this point of the project it is capable of importing the schema of any relational database for which there is a JDBC driver), augment the schema by defining new inter-database dependencies (foreign keys), define where the social-demographic data about the patients is located in order to allow the matching of patient identifiers and define which data from the underlying database are shared.
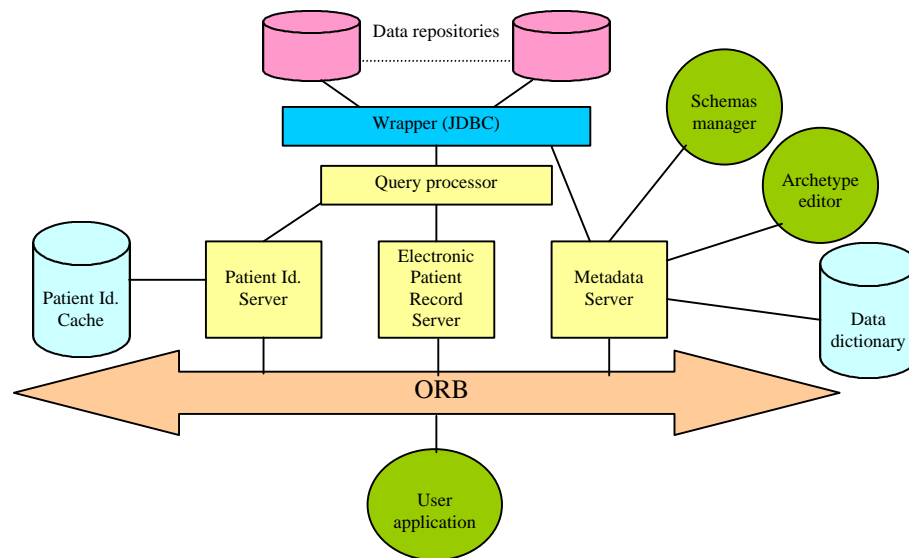
Fig2. System architecture to support integration of distributed health data

- The Patient Identification service allows the identification of identical patients across different system when there is not a universal patient ID. For this purpose a set of social-demographic attributes are used (SSN, surnames, national identification number, date of birth, etc.). The search begins by matching the supplied attributes exactly (e.g. surnames, national identification number and date of birth). If none is found (for instance due to spelling errors), the condition is weakened and all the patients that match in a random part of the supplied attributes are selected, the Levenstein distance algorithm (LDA) is used to select and sort the most similar ones. Finally, the users select the correct one and the matched patient IDs are stored in a cache database to be reused later on.

- The Electronic Patient Record Server is the core of the whole system. It is layered between the client applications and the data repositories. This server retrieves, by request, all the relevant patient information wherever it is located and presents back the information in a uniform way to the user applications. User applications request to the EPR service healthcare information about a particular patient as one or more instances of any archetype defined in the data dictionary. The EPR service obtains the definition and mappings of the requested archetype from the metadata service. Afterwards, it builds and populates (by executing the mappings) the objects that contain the health data.

V. CONCLUSIONS

Healthcare is rapidly taking on a distributed nature, thus the ability to share effectively, meaningfully and securely health data about patients is the key issue in providing good and cost-effective healthcare. The above outlined system model and architecture used to define a server capable of integrated faithfully distributed patient's healthcare information across an institution. The EHCR architecture used is ENV 13606 from CEN/TC 251, however some extension has been developed to cope with the problem of data distribution among several pre-existing information systems. Our solution is based on defining a set of formalised aggregates of data with specific semantics and associating them with the heterogeneous structures found in the autonomous information systems. This approach is similar to others found in the literature [5][6]. The overall server comprises several servers interconnected by CORBA. The work described in this paper is still in progress, currently the development of the project is going according to plan and Hospital Lluis Alcanys of Xàtiva (Valencia) is being used as validation site. This project is the first step towards the achievement of some kind of electronic healthcare record for the hospitals of the Valencian healthcare network.

REFERENCES

1. Sheth, A. P.; Larson, J.A.: Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys, 22-3 (1990) 183-235.
2. Jung, B., Grimson, W., Crimson, J.: The EHCR-if not now, when?. Toward an Electronic Health Record Europe' 99, Proceedings Book (1999), 51-55.
3. Batini, C., Lenzerini, M. Navathe. S.B.: A comparative analysis of methodologies for database schema integration. ACM computing surveys, 18-4 (1986) 323-364.
4. CEN/TC251 WG I.: Health Informatics-Electronic Healthcare Record Communication- Part 1: Extended architecture and domain model, Final Draft prENV13606-1 (1999).
5. Thurlet, G, et al.: ARCHIMED: A network of integrated information systems. Methods of Information in Medicine (2000), 36-43.
6. Grimson, J., et al.: a CORBA-Based integration of distributed electronic healthcare records using the Synapses approach. IEEE transactions on information technology in biomedicine, 3-4 (1998), 124-137.